Pattern Recognition 46 (2013) 3533–3547

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Learning small gallery size for prediction of recognition performance on large populations

Rong Wang, Bir Bhanu, Ninad S. Thakoor *

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 30 October 2012
Received in revised form
8 May 2013
Accepted 20 May 2013
Available online 7 June 2013

Keywords:
Biometrics
Distortion modeling
Learning
Optimal small gallery size
Performance bounds
Performance prediction

ABSTRACT

This paper addresses the estimation of a small gallery size that can generate the optimal error estimate and its confidence on a large population (relative to the size of the gallery) which is one of the fundamental problems encountered in performance prediction for object recognition. It uses a generalized two-dimensional prediction model that combines a hypergeometric probability distribution model with a binomial model and also considers the data distortion problem in large populations. Learning is incorporated in the prediction process in order to find the optimal small gallery size and to improve the prediction. The Chernoff and Chebychev inequalities are used as a guide to obtain the small gallery size. During the prediction, the expectation–maximization (EM) algorithm is used to learn the match score and the non-match score distributions that are represented as a mixture of Gaussians. The optimal size of the small gallery is learned by comparing it with the sizes obtained by the statistical approaches and at the same time the upper and lower bounds for the prediction on large populations are obtained. Results for the prediction are presented for the NIST-4 fingerprint database.

1. Introduction

Recognition systems can classify images, signals, or other types of measurements into a number of classes. In this paper, we mainly focus on biometrics recognition systems. Biometrics can be a fingerprint, a palmprint, a face image, gait, signature, speech, etc. Depending on the application there are two kinds of biometric recognition systems: verification systems and identification systems. Verification (also called authentication) is a one-to-one matching problem [1]. A verification system stores users' biometrics in a database. Then, it compares a person's biometrics signatures with the stored representation to verify if this person is indeed who she/he claims to be. The system can accept or reject the claim according to the verification result. An identification system is more complex than a verification system. In an identification system, for a given query, the system searches the entire database to find out if there are any biometrics signatures that match the query. It conducts a one-to-many matching. There are two kinds of identification systems: the closed-set identification systems and the open-set identification systems [2]. The closed-set identification is the identification for which all potential users are enrolled in the system. Alternatively, the open-set identification is the identification for which some potential users are not enrolled in the system. The verification and the closed-set identification can be considered to be special cases of the open-set identification.

In a practical recognition system, some important parameters for characterizing the system are generally unknown [3]. We need to predict these parameters from a set of available data. In this paper, we provide a prediction model for performance of a closed-set identification system. Since the recognition performance of an algorithm is usually estimated based on limited data, it is difficult to predict its performance for additional data: the limited test data may, after all, not accurately represent a larger population. Before we can evaluate and predict the performance of a recognition algorithm on large populations, we need to answer some fundamental questions. When we use a small gallery to estimate the algorithm performance on large populations, how can we find the optimal size of the small gallery and how accurate is the estimation? Since the prediction is based on the same recognition algorithm, we can give the confidence interval for the performance estimation on a large population [4]. The confidence interval [5] can describe the uncertainty associated with the estimation. This gives an interval within which the true performance of the algorithm for a large population is expected to fall, along with the probability that it is expected to fall there [6]. Bolle et al. [7] presented a bootstrap based approach to compute the confidence interval to evaluate the biometrics system performance.

In this paper, we address the problems associated with the prediction of performance on large populations and the optimal

* Corresponding author. Tel.: +1 951 827 3954.
E-mail addresses: bhanu@cris.ucr.edu (B. Bhanu), ninadt@ucr.edu (N.S. Thakoor).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.05.024

small gallery size. The term small gallery is used to emphasize the size of the gallery used during the design of a biometric system which is small compared to the population. We use a generalized prediction model for a closed-set identification system that combines a hypergeometric probability distribution model with a binomial model. Hypergeometric distribution is a discrete probability distribution which captures the probability of picking a certain number of good samples from a mix of good and bad samples without replacement.

The prediction model takes into account distortions that may occur in large populations. When a physical phenomenon is observed and a quantity corresponding to its properties is measured, the measurement differs from the true underlaying value. This discrepancy is called the distortion. The model also provides performance measurements as a function of the rank, the large population size, the number of distorted images, and match and non-match score distributions.

We model the match score and the non-match score distributions as mixture of Gaussians and use the expectation–maximization (EM) algorithm to estimate its parameters. Given limited data, we can use parametric or nonparametric estimation methods to estimate the data distribution. The expectation–maximization (EM) algorithm [8], one of the parameter estimation methods, assumes that the underlying distribution is known. It is an iterative method to estimate the mixture parameters by maximum likelihood techniques. We introduce learning by feeding back the similarity scores (match scores and non-match scores) to increase the small gallery size. In this way, we can find the optimal size of the small gallery to predict the large population performance.

We also provide the upper and the lower bounds for the prediction performance of a large population. We use two different statistical methods—Chernoff's inequality and Chebychev's inequality—to obtain the relationship between the small gallery size and the confidence interval for a given margin of error. In probability theory, inequalities such as Chernoff's and Chebyshev's are routinely used to provide bounds on the distribution values when minimal information (e.g. mean and standard deviation for Chebyshev's) regarding the distributions is available.

The specific contributions of the paper are:

(1) We use a generalized prediction model that combines a hypergeometric probability distribution model with a binomial model which takes into account distortions that may occur in large populations. Our distortion model includes feature uncertainty, feature occlusion, and feature clutter. In the prediction model, we model the match score and non-match score distributions as a mixture of Gaussians, use the EM algorithm to estimate its parameters and find the number of components of the distributions automatically.
(2) We find the optimal size of a small gallery by an iterative learning process. We use the Chernoff inequality and the Chebychev inequality to determine the small gallery size in theory which is related to the margin of error and the confidence interval. We find the upper bound and a good lower bound on recognition performance on a large population.
(3) Systematic experimental results are shown on a challenging large data set of fingerprint images (NIST-4) with realistic distortion models.

The paper is organized as follows. Related work is presented in Section 2. The details of the technical approach are given in Section 3. It includes the distortion model, the prediction model, and the statistical methods to find the relationship between the optimal small gallery size and the confidence interval. Experimental results are provide in Section 4. The combined model with learning is tested on the *NIST Special Database 4* (NIST-4) which is the rolled fingerprint database. Conclusions are presented in Section 5.

2. Related work

Many researchers have used statistical approaches to estimate the performance of recognition systems. Usually, these approaches use prediction models based on the feature space or similarity scores. Wayman [9] and Daugman [10] developed a binomial model that used the non-match score distribution. This model underestimates recognition performance for large populations [11]. Phillips et al. [12] developed a moment model, which used both the match score and non-match score distributions.

Pankanti et al. [13] presented a fingerprint individuality model which was based on the feature space and derived an expression to estimate the probability of false matching between two fingerprints based on minutiae. The model measured the amount of information needed to establish correspondence between two fingerprints. Tan and Bhanu [14] presented an improvement over [13] by providing a two-point model and a three-point model to estimate the error rate for the minutiae based fingerprint recognition. Their approach measured minutiae's position and orientation and the relations between different minutiae to find the probability of correspondence between fingerprints. They allowed overlap of the uncertainty area of any two minutiae.

Johnson et al. [15] improved the moment model by using a multiple non-match score set. They averaged match scores of the entire gallery. For each match score, they counted the number of non-match scores larger than the match score leading to an error. They assumed that the match scores are distributed uniformly. Grother and Phillips [11] introduced a joint density function of the match score and the non-match score to estimate both the open-set and the closed-set identification performance. Since the joint density is generally impractical to estimate, they assumed that the match score and non-match scores are independent and their distributions are the same for large populations. They used the Monte Carlo sampling method to linearly interpolate the match score and the non-match score look-up tables. Tabassi et al. [16] and Wein and Baveja [17] used the fingerprint image quality to predict the performance. They defined the quality as an indication of the degree of separation between the match score and non-match score distributions. The farther these two distributions are, the better the system performs.

Ju and Bhanu [18] predicted the gait recognition performance by probabilistic simulation of different within-class feature variance. They provided the upper bound for the recognition performance with regard to different human silhouette resolutions. Li et al. [19] developed an analytical performance characteristic to predict the misclassification statistics of the resulting boosted classifier. The analytic error characterization establishes the relationship between the misclassification statistics and the size of training set and the true distribution parameters.

Wang et al. [20] trained a support vector machine from features based on match and non-match scores to predict success and failure of the face recognition. Scheirer et al. [21] analyzed similarity surfaces to predict algorithmic failures in face recognition for various face recognition algorithms. Aggarwal et al. [22] learned mapping from the image characterization space to the score space to predict performance of face recognition algorithms on unseen data.

Usually, for a biometrics recognition system, the performance margin of error is prespecified. Consequently, providing the upper and lower bounds for the performance is another important topic in the recognition performance prediction. Lindenbaum [23]